

# RuThes Cloud: Towards a Multilevel Linguistic Linked Open Data Resource for Russian

Alexander Kirillovich<sup>1</sup>(✉), Olga Nevzorova<sup>2</sup>, Emil Gimadiev<sup>1</sup>,  
and Natalia Loukachevitch<sup>3</sup>

<sup>1</sup> Kazan Federal University, Kazan, Russia

al.k.kirillovich@gmail.com, gimpw1@gmail.com

<sup>2</sup> Institute of Applied Semiotics, Tatarstan Academy of Sciences, Kazan, Russia

onevzoro@gmail.com

<sup>3</sup> Lomonosov Moscow State University, Moscow, Russia

louk\_nat@mail.ru

**Abstract.** In this paper we present a new multi-level Linguistic Linked Open Data resource for Russian. It covers four linguistic levels: semantic, lexical, morphological and syntactic. The resource has been constructed on base of the well-known RuThes thesaurus and the original hitherto unpublished Extended Zaliznyak grammatical dictionary. The resource is represented in terms of SKOS, Lemon, and LexInfo ontologies and a new custom ontology. Building the resource, we automatically completed the following tasks: merging source resources upon common lexical entries, decomposing complex lexical entries, and publishing constructed resource as LLOD-compatible dataset. We demonstrate the use case in which the developed resource is exploited in IR task. We hope that our work can serve as a crystallization point of the LLOD cloud in Russian.

**Keywords:** Linguistic Linked Open Data · Linked data · Language resources · Ontology · Thesaurus · Lexicon · Grammatical dictionary · Dependency grammar · RuThes · Russian language

## 1 Introduction

Publishing linguistic data according to the principles of Linked Open Data is a recently established trend, that has led to the emergence of the fast-growing Linguistic Linked Open Data (LLOD) cloud [1, 2]. It has the following advantages:

- Structural and conceptual interoperability.
- Integration of linguistic resources (for example, of a corpus and a thesaurus linked to it) to solve a common problem.
- Integration of linguistic and extralinguistic resources from the Linked Open Data cloud. This integration is exploited in such tasks as: (1) Ontology-based information extraction from text. (2) Semantic annotation and search for documents on the LOD cloud. (3) Generation of SPARQL queries for user-provided natural language questions. (4) Production of human understandable descriptions of LOD resources.

© Springer International Publishing AG 2017

P. Rózewski and C. Lange (Eds.): KESW 2017, CCIS 786, pp. 38–52, 2017.

[https://doi.org/10.1007/978-3-319-69548-8\\_4](https://doi.org/10.1007/978-3-319-69548-8_4)